Interactions Between Genome Architecture and Virulence Genes in *Pseudomonas syringae,* strain CC1557 as a model

Kevin L. Hockett[1☯], Marc T. Nishimura[2☯], Erik Karlsrud[1], Kevin Dougherty[1], and David A. Baltrus[1*]

[1]School of Plant Sciences, University of Arizona, Tucson AZ, 85721-0036
[2]Department of Biology, University of North Carolina, Chapel Hill, NC 27599
☯ these authors contributed equally

*Corresponding author:

Baltrus@email.arizona.edu
The University of Arizona
School of Plant Sciences
P.O. Box 210036
Forbes Building, 303
Tucson, Arizona, USA 85721-0036
Phone: (520) 626-8215
Fax: (520) 621-7186

Abstract

Both type III effector proteins and non-ribosomal peptide toxins play important roles for *Pseudomonas syringae* pathogenicity in host plants, but if and how these virulence pathways interact to promote infection remains unclear. Genomic evidence from one clade of *P. syringae* suggests a tradeoff between the total number of type III effector proteins and presence of syringomycin, syringopeptin, and syringolin A toxins. Here we report the complete genome sequence from *P. syringae* CC1557, which contains the lowest number of known type III effectors to date and has also acquired a highly similar toxin to syringomycin through horizontal transfer. We demonstrate that this strain is pathogenic on *Nicotiana benthamiana* and that both the type III secretion system and a new type III effector family, *hopBJ1*, contribute to virulence. Taken together, our results provide independent evolutionary confirmation of a negative correlation between type III effector repertoires and some non-ribosomal peptide toxins while also highlighting how genomic synteny can be used to identify novel virulence proteins.

Introduction

Pseudomonas syringae, a bacterial phytopathogen of many crop species, utilizes a diverse arsenal of virulence factors in order to infect host plants (Hirano and Upper, 2000; O'Brien et al., 2011). Much of the previous research on pathogenicity in *P. syringae* has focused on identifying and characterizing novel virulence factors within strains and on categorizing their presence across the species (Baltrus et al., 2011;

Hwang et al., 2005; Lindeberg et al., 2012; O'Brien et al., 2011). Although this accumulation of knowledge has enabled tests of protein function during infection, many questions remain as to how virulence pathways interact and evolve at a systems level (Baltrus et al., 2011; Cunnac et al., 2011; Hwang et al., 2005). In the absence of direct tests, identification of similar genomic trends across independent lineages provides additional support for the presence of such interactions over evolutionary time scales.

One of the main contributors to virulence within *P. syringae* is a type III secretion system (T3SS) (O'Brien et al., 2011). The T3SS encodes a structure that translocates bacterial effector proteins (T3E) into host cells in order to disrupt host physiological pathways and enable successful infection (Lindeberg et al., 2012). *P. syringae* genomes typically contain about 10-40 total T3Es, the exact actions of which depend on the totality of effector repertoires as well as host genotype (Baltrus et al., 2011; Lindeberg et al., 2012; O'Brien et al., 2011). Another main contributor to pathogenesis within many *P. syringae* strains are non-ribosomal peptide (NRP) toxin pathways. NRP pathways encode proteins that act as an assembly line, elongating and decorating peptide chains involved in pathogenicity (Bender et al., 1999). NRP pathways are often not conserved between closely related strains, and even if present, can differ in their outputs based on nucleotide polymorphisms (Baltrus et al., 2011; Bender et al., 1999; Hwang et al., 2005).

Multiple reports have suggested functional overlap or phenotypic interactions between T3E and NRP pathways. For instance, phenotypic virulence functions of the toxin coronatine can be partially restored by the toxin syringolin A as well as T3Es

HopZ1 and AvrB1 (Cui et al., 2010; Jiang et al., 2013; Melotto et al., 2006; Schellenberg et al., 2010). Furthermore, within MLSA group II *P. syringae* strains as defined by Sarkar and Guttman (2004), there exists a negative correlation between presence of conserved toxin pathways (syringolin A, syringopeptin, and syringomycin) and size of the T3E repertoire compared to other phylogenetic clades (Baltrus et al., 2011). In fact, MLSA group II strains possess the lowest number of T3E among analyzed genomes across the species. This correlation is strengthened by the observation that strains from pathovar *pisi* possess the largest T3E repertoire within MLST group II and have lost all three toxin pathways (Baltrus et al., *in press*). Whether this negative correlation between NRPs and T3Es reflects an unrecognized difference in disease ecology is yet to be determined, although strains from MLST group II are thought to survive as epiphytes much better than other clades within *P. syringae* (Feil et al., 2005).

The hypothesis of a phenotypic tradeoff between T3E and NRP pathways would be bolstered by discovery of an independent clade of *P. syringae* that has gained similar toxins to the MLST group II strains but which has also lost effectors. We have analyzed the genomes of a small clade of *P. syringae* isolated as non-pathogens from environmental sources, and have found independent evidence supporting this genomic trend. These strains have lost T3Es relative to closely related strains, contain the smallest number of known T3E families to date, and appear to have acquired at least one NRP pathway with similarity to genes encoding syringomycin. We demonstrate that one of these strains, CC1557, can infect *Nicotiana benthamiana* and cause T3SS-dependent disease. We further show that virulence of this strain is significantly

increased by the presence of one new T3E effector family, HopBJ1, which is similar to cytotoxins found within other bacteria. Therefore, independent evidence suggests that acquisition NRP pathways correlates strongly with the loss of T3E families and further strengthens the idea of an evolutionary tradeoff between these virulence factors.

Results

A complete genome sequence for *P. syringae* CC1557

Using a combination of 100bp Illumina paired end reads and longer PacBio reads, we assembled a complete genome sequence for strain CC1557. This genome consists of a 5,728,024 bp chromosome and a 53,629 bp plasmid. According to RAST annotation, this chromosome contains 5119 ORFs while the single plasmid contains an additional 68 ORFs. (We are currently in the seemingly never-ending process of revising the genome sequence within GenBank, but the full sequence sequence can be found at https://dl.dropboxusercontent.com/u/7891644/ CC1557_fullchromosome_dnaA_withplasmid.fa)

P. syringae CC1557 can infect and cause disease in *N. benthamiana*

P. syringae CC1557 was originally isolated from snow, while the closely related strain CC1466 was originally isolated as an asymptomatic inhabitant of *Dodecantheon pulchellum*, a perennial herb (Morris et al., 2008). Using syringe inoculation under standard conditions, we have demonstrated that *P. syringae* CC1557 can grow to high levels in the apoplast of *N. benthamiana* after 3 days of infection (Figure 1).

Furthermore, these large bacterial populations cause disease symptoms in that tissue collapse is visible within 3 days of syringe inoculation (Figure 2). High levels of growth and tissue collapse are both eliminated by deleting the *hrcC* gene (Figure 2) from CC1557, which encodes a main structural protein for the type III secretion pilus. Therefore, CC1557 virulence in *N. benthamiana* under laboratory conditions is T3SS dependent.

The genomes of *P. syringae* CC1557 and CC1466 encode a reduced type III effector repertoire. We used bioinformatic methods as per Baltrus et al. 2011, to search the draft genomes of *P. syringae* strains CC1557, CC1466, and the closely related strain CC1583, for known type III effector families (Baltrus et al. *submitted*). We found a total of 4 T3E shared across these three genomes, with 8 T3E found only within CC1583 (Table 2). Therefore, together with HopBJ1(see below), the genomes of CC1466 and CC1557 appear to encode a total of 5 potential T3E from known protein families. This stands in contrast to the immediate outgroup strain CC1583, which encodes a total of 12 potential T3E from known protein families.

***P. syringae* CC1466 and CC1557 have horizontally acquired pathways for production of non-ribosomal peptides that resemble syringopeptin and syringomycin.** We used tBLASTn searches of the complete genome sequence of CC1557 in order to identify potential NRP toxin pathways as per Baltrus et al. 2011. This genome contains highly similar full length tBLASTn matches to a variety of proteins involved in syringomycin synthesis: SyrB (93% Identity), SyrC (99%), and SyrP (93%),

data not shown. Further investigation of all complete genomes for *P. syringae* demonstrates that the region containing putative NRP-related loci in CC1557 is found in approximately the same genomic context on the chromosome as the syringomycin-syringopeptin pathways in *P. syringae* pv. *syringae* B728a (Figure S1). Moreover, the genomic context surrounding these toxin pathways is conserved throughout *P. syringae,* except that the region of CC1557 is inverted compared to other genomes (Figure S1). Although regions of high similarity to the syringopeptin pathway do not appear to be present within CC1557, analysis of gene annotations suggests that a second NRP pathway occurs immediately downstream of the syringomycin-like region on the chromosome of CC1557 (data not shown). While it is difficult to identify the NRP product of these CC1557 pathways by sequence alone, draft genomes sequences suggest that this region is present in CC1466 but not CC1583. The most parsimonious explanation for this phylogenetic signal is that an immediate ancestor of CC1466 and CC1557 acquired this NRP region by horizontal transfer after divergence from the CC1583 lineage.

A new type III effector protein family is present in the genomes of both *P. syringae* CC1466 and CC1557. Similarity searches of the *P. syringae* genomes demonstrated that *hopM1*, which encodes an effector protein largely conserved throughout *P. syringae*, had likely been lost within strains CC1466 and CC1557. Further investigation demonstrated that an unknown protein open reading frame had replaced *hopM1* within these genomes (Figure 3). In order to test if this unknown protein is translocated into plant cells, we cloned it into the translocation test vector pBAV178 and tested for

delivery of avrRpt2 as a fusion protein. We found that the resulting HopBJ1-avrRPT2 fusion protein triggered cell death in both *Arabidopsis thaliana* Col-0 wild type and *rps2* lines, indicating that HopBJ1 itself was capable of triggering cell death in the absence of recognition of avrRpt2 (Table 3). In order to verify that HopBJ1 triggered cell death without avrRpt2 we delivered HopBJ1:HA into a variety of plant genotypes using either *Pto* DC3000. Delivery of a HopBJ1:HA can cause tissue collapse in a variety of Arabidopsis accessions (Table 3). All genotypes tested displayed strong cell death by either 24 or 48hr post inoculation.

HopBJ1 cell death is not dependent on the known R-gene pathway mutants *eds1*, *pad4* or *rar1* (Table 3). HopBJ1-induced cell death also occurs in plants expressing the salicylic acid degrading enzyme NahG (Table 3). However, this rapid tissue collapse fails to occur out of a *hrcC-* background of *Pto* DC3000, indicating HopBJ1 secretion is dependent on a functioning T3SS (data not shown). The genomic region immediately upstream of *hopBJ1* possesses a canonical hrp-box sequence (Fig. S2), and so regulation of this gene very likely takes place through the action of the sigma factor HrpL. We therefore conclude that this novel ORF encodes a new putative T3E family, HopBJ, and that this family possesses broad cytotoxic capabilities.

Deletion of HopBJ1 from *P. syringae* CC1557 causes loss of disease symptoms and lowers bacterial populations *in planta*. We have created a deletion of *hopBJ1* in *P. syringae* strain CC1557, and have performed growth curves *in planta* using syringe infiltration. Deletion of *hopBJ1* causes a repeatable loss of growth *in planta* after two

days of infection (Figure 4). We further find that deletion of *hopBJ1* leads to a loss of the tissue collapse phenotype after 3 days (Figure 2). We have been able to complement both of these phenotypes *in planta* through expression of *hopBJ1* using its native promoter *in trans* (Figures 2 and 4).

HopBJ1 resembles cytotoxic genes present within other species. Currently the closest blast hits for HopBJ1 are hypothetical proteins from *Serratia marcescens* and *Hahella chejuensis*. However, modeling of protein structure using the Phyre2 web server demonstrates that HopBJ displays limited similarity to the *E. coli* CNF1 toxin (Figure 5). Surprisingly, amino acids critical to the deamidase functions of CNF1 appear to be conserved in HopBJ1 (Figure 5) as well as in similar proteins from *S. marcescens* and *H. chejuensis* (data not shown). This modeling of protein structure further suggests that C174 and H192 are a functional catalytic dyad within HopBJ1.

Discussion

Recent advances in genome sequencing have spurred great interest in conservation and diversification of virulence pathways across *Pseudomonas syringae* (O'Brien et al., 2011; Studholme, 2011). While genome gazing often leads to the identification of trends across a species, single observations must be treated with skepticism in the absence of additional data. One particularly interesting, yet relatively little understood, pattern across genomes involves a negative correlation in MLSA group II between size of T3E repertoire and presence of NRP pathways (Baltrus et al., 2011).

Importantly, this trend does not appear to be a result of sampling bias or failure to discover novel T3E because strains within this clade have been thoroughly screened at both genetic and transcriptome level. Indeed the entire *hrpL* regulon, the major transcriptional regulator for all *P. syringae* T3E, is reduced within MLST group II compared to other clades (Mucyn et al., *in press*).

Strains CC1466 and CC1557 diverge early in the phylogeny of *P. syringae sensu latu*, and have not been isolated from diseased plants in nature (Morris et al., 2008). We have shown that CC1557 can act as a pathogen of *N. benthamiana* and, as with all other phytopathogenic *P.syringae* strains, that virulence requires a functional T3SS. Surprisingly, we have found that both of these genomes contain few T3Es, with only 5 known loci present and shared by both. Genomic comparisons between these two strains have further demonstrated that both contain an NRP toxin pathway very similar to those encoding syringomycin. As is the nature of NRP toxins even a single amino acid change can have dramatic consequences on the final toxin product chemistry (Bender et al., 1999), but it is important to note that these many of the potential ORFs in these regions are highly similar (>90% protein ID) to syringomycin ORFs present within MLST group II strains. Comparison to closely related outgroup strains demonstrates that acquisition of this NRP pathway is evolutionarily correlated with loss of T3E families. Indeed, the most closely related sequenced outgroup strain, CC1583, lacks this NRP pathway and contains twice the number of known T3E loci (Table 2). As such, we interpret these patterns as independent evidence suggesting an evolutionary or ecological tradeoff between presence of syringomycin and number of T3Es within the

strain. We also note that there appears to be an additional, uncharacterized NRP pathway downstream of the syringomycin-like region within CC1557 (data not shown).

What selective pressure could underlie such a negative correlation between T3E and NRP toxins? It is possible that the functions of these NRP toxins are redundant with a specific suite of T3Es. In this case, T3E loss can take place because there is no longer positive selection to counter selection against evolutionary (such as recognition by the plant immune system) or physiological costs. For MLSA clade II and CC1557, acquisition of syringomycin and potentially additional NRP pathways could render the functions of some T3E obsolete. If this is the case, we further predict that the potential second NRP toxin pathway (downstream of the syringomycin-like pathway in CC1557) could have similar functions as syringopeptin. It's also possible that disease ecologies of most MLST group II strains, as well as CC1466 and CC1557, differ compared to other focal *P. syringae* strains. We have a rudimentary understanding of differences in disease ecology throughout *P. syringae*, although it is recognized that differences do exist across the phylogeny (Clarke et al., 2010). For instance, infection of woody hosts has arisen multiple independent times on multiple hosts yet such strains are still dependent on a functioning T3SS (O'Brien et al., 2011; Ramos et al., 2012). Strains that contain syringopeptin and syringomycin potentially cause disease on a wider host range than other groups, which may be enabled by generality of NRP toxins (Baltrus et al., 2011; Hwang et al., 2005; Quigley and Gross, 1994). MLST group II strains appear to survive better as epiphytes compared to other groups, and presence of NRP toxins may tie into this strategy (Feil et al., 2005). Moreover, such differences in disease

ecology may be difficult to precisely measure because they may only be visible under natural conditions of infection and dispersal.

It is possible that there are a number of undescribed T3E families within CC1466 and CC1557, and to this point we have successfully used comparisons of genome synteny to identify HopBJ1. *hopBJ1* is present within the conserved effector locus within these strains, and as such is proximate to both the structural genes for the T3SS and the T3E *avrE*. While it is difficult to precisely map out the evolutionary history of this region, it does appear that *hopBJ1* has been recombined into this locus in place of *hopM1*. In some respects, recombination of *hopM1* from this locus is analogous to a situation witnessed in MLST group I *P. syringae* strains where *hopM1* alleles have been cleanly swapped through recombination (Baltrus et al., 2011). HopM1 is an effector whose presence (although not necessarily in functional form) is conserved throughout *P. syringae sensu strictu*, and has been shown to act redundantly with AvrE during infection of plant hosts (Badel et al., 2006; Kvitko et al., 2009). It is currently unknown whether HopBJ1 can act redundantly AvrE on specific hosts, although that our single deletion mutant shows a virulence phenotype on *N. benthamiana* speaks against this possibility. Moreover, since HopBJ1 appears to be responsible for a significant portion the growth of CC1557 *in planta,* it will be interesting to see how this effector behaves within different strain backgrounds on different hosts.

HopBJ1 itself displays protein structure similarity to CNF1 toxin found within pathogenic *E. coli* strains. CNF1 functions by causing deamidation of a glutamine

residue, crucial for GTP hydrolysis for small GTPases of the Rho family, thereby leading to constitutive activation and actin disruption (Lemonnier et al., 2007). The pathology of HopBJ1 on Arabidopsis, when delivered from strain *Pto* DC3000, suggests that HopBJ1 could be acting as a general toxin because cell death appears regardless of accession and is not dependent on typical R-gene related host defense pathways (Table 3). In some ways, this may be similar to the functions of AvrE, which was identified as causing cell death outside of an R-gene dependent HR response. It is tempting to think that both AvrE and HopBJ1 ultimately lead to the same physiological outcomes during infection, but so far little is known about processes targeted by AvrE during infection. What advantages this protein provides to CC1557 in nature remains unknown, but an intriguing possibility is that the disease ecology of CC1557 has changed and that this strain is now more necrotrophic than other *P. syringae* strains.

Herein we have analyzed the genomes of two *P. syringae* strains isolated from environmental sources. These genomes contain a paucity of known T3E, even though virulence of CC1557 on *N. benthamiana* is dependent on a functioning T3SS. That these strains have also acquired a NRP pathway similar to syringomycin represents an independent evolutionary example that supports a negative correlation between NRP pathways and T3E repertoires. Furthermore, this data set highlights how genomic scans can lead to insights into pathogenesis while also demonstrating the power of genetic diversity to uncover genome-wide changes in virulence gene architecture.

Methods

Plasmids, bacterial isolates, and growth conditions.

All bacterial strains and plasmids used or created are listed in table 1. Typically, *P. syringae* isolates were grown at 27°C on KB media using 50**µ**g/mL rifampicin. When necessary, cultures of both *P. syringae* and *E. coli* were supplemented with antibiotics or sugars in the following concentrations: 10**µ**g/mL tetracycline, 50**µ**g/mL kanamycin, 25**µ**g/mL gentamycin, and 5% sucrose.

All clones were created by first amplifying target sequences using *Pfx* polymerase (Invitrogen), followed by recombination of these fragments into the entry vector pDONR207 using BP clonase (Invitrogen). All ORF (without a *hrp*-box promoter) and gene (including the promoter) sequences were confirmed by Sanger sequencing of these pDONR207 clones. Clones in entry vectors were recombined into destination vectors of interest using LR clonase (Invitrogen).

Genome Sequences and Searches

Draft genome assemblies for CC1466, CC1557,and CC1583 are publicly available through Genbank (accession numbers AVEM00000000, AVEH00000000, and AVEG00000000 respectively; Baltrus et al., *submitted*). One PacBio SMRT cell yielded 37,509 reads for a total of 268,122,626 nucleotides after filtering for quality. To piece together the complete genome sequence for CC1557, PacBio reads were first

assembled using the HGAP software (Chin et al., 2013), yielding two contigs total with no scaffolding gaps. Contigs from the previous assembly (using only 100bp Illumina paired end reads, Baltrus et al. *submitted*) were then overlayed onto this HGAP assembly. When there was coverage by contigs from the Illumina assembly, the previous assembly sequence was chosen as the final sequence, but when there was no coverage the PacBio assembly was chosen. In cases where multiple divergent Illumina contigs assembled to the same region in the HGAP assembly, the Illumina sequence which matched the PacBio assembly was chosen. Since the second contig from the HGAP assembly possesses numerous plasmid related genes, we are confident that this does actually represent a plasmid present within this strain. Both the chromosome and plasmid were annotated using the RAST server (Aziz et al., 2008).

T3Es and NRP pathways were identified as per Baltrus et al. (2011). Briefly, draft genome assemblies were queried using tBLASTn with queries consisting of protein sequences for known type III effector families or for key loci within toxin pathways. Each BLAST hit was validated by hand for copy number, identity, and completeness. Briefly, all BLAST results were visually inspected to make sure that each sequence displayed high similarity to only one region in the assembly, that this similar region was not part of a larger ORF, and that the length of this region was greater than 40% of the length of the original query sequence.

The Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) was used to carry out protein structure similarity searches (Kelley and Sternberg, 2009).

Generation of mutants

Bacterial mutants were generated as per Baltrus et al. 2012. Regions (>500bp) upstream and downstream of the target genes were PCR amplified separately and then combined into one fragment by overlap extension PCR. The bridge PCR amplicon was then cloned into pDONR207, and further moved into pMTN1907 using LR clonase. Once mated into *P. syringae,* single recombination of a homologous region upstream or downstream of the target region and subsequent selection on sucrose allows for screening of clean deletions. Mutants were confirmed by phenotyping for sucrose resistance, tetracycline sensitivity, PCR amplification of the deletion, and failure of PCR to amplify regions within the deletion.

In planta *Growth Curves and Translocation Tests*

Nicotiana benthamiana plants were grown for 4-6 weeks on a long day cycle (16 hours light/8 hours dark). Plants were removed from the growth chamber and allowed to acclimate to ambient laboratory conditions for a period of 3-5 days prior to infiltration. Bacteria were cultured over night on KB with amended with proper antibiotics. Bacterial cells were washed 1x in and resuspended to an OD_{600} of 0.002 in sterile 10 mM $MgCl_2$ yielding a final inoculum density of approximately $1x10^6$ CFU/mL. Cell suspensions were then syringe infiltrated into the abaxial side of *N. benthamiana* leaves. Populations were recovered after 3 days of growth using a

corkborer. Leaf disks were disrupted using glass beads and a bead-beater device and populations were enumerated by dilution plating onto KB amended with appropriate antibiotics.

HopBJ1 was tested for the ability to secrete the active C-terminal fragment of AvrRpt2 to cause a hypersensitive response (HR) in Arabidopsis accession Col-0 (Guttman et al., 2002). For this test, promoterless *hopBJ1* was cloned into plasmid pBAV178 (Vinatzer et al., 2006) so that expression was dependent on a *tet* promoter from strain *Pto* DC3000. This construct was also placed into a *hrcC-* mutant of *Pto* DC3000 (Baltrus et al., 2012), in order to test the requirement of a functioning TTSS for tissue collapse. All HR tests were performed on ~5 week old plants and utilized a bacterial density of OD_{600} 0.05. Tissue collapse was measured at either 24 or 48 hours post infection and confirmed by at least four independent tests.

Phylogenetic Methods

Phylogenetic analyses were performed on concatenated MLSA loci as described previously using concatenated fragments from 7 MLSA genes across a diverse array of isolates from P. syringae (Baltrus et al, 2011) as well as AvrE and HrpW protein sequences. MrBayes was used to perform Bayesian phylogenetic analyses with flat priors, a burn-in period of 250,000 generations, and convergence after 1,000,000 total generations (Ronquist et al., 2012).

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., et al. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics *9*, 75.

Badel, J.L., Shimizu, R., Oh, H.-S., and Collmer, A. (2006). A *Pseudomonas syringae* pv. *tomato avrE1/hopM1* mutant is severely reduced in growth and lesion formation in tomato. Mol Plant Microbe Interact *19*, 99–111.

Baltrus, D.A., Nishimura, M.T., Romanchuk, A., Chang, J.H., Mukhtar, M.S., Cherkis, K., Roach, J., Grant, S.R., Jones, C.D., and Dangl, J.L. (2011). Dynamic Evolution of Pathogenicity Revealed by Sequencing and Comparative Genomics of 19 *Pseudomonas syringae* Isolates. PLoS Pathog *7*, e1002132.

Baltrus, D. A., Nishimura, M. T., Dougherty, K. M., Biswas, S., Mukhtar, M. S., Vicente, J.*, et al.* (2012) The molecular basis of host specialization in bean pathovars of *Pseudomonas syringae*. Mol Plant Micbobe Interact **25,** 877-888.

Baltrus, D.A., Dougherty, K., Beckstrom-Sternberg, S.M., Beckstrom-Sternberg, J.S., and Foster, J.T. (*in press*) Incongruence between MLSA and whole genome phylogenies: *Pseudomonas syringae* as a cautionary tale. Mol Plant Pathol.

Baltrus, D.A., Yourstone, S., Lind, A., Guilbaud, C., Sands, D.C., Jones, C.D., Morris, C.E., Dangl, J.L. (*submitted*). Draft Genomes for a phylogenetically diverse suite of *Pseudomonas syringae* strins from multiple source populations. *Genome Announcements.*

Bender, C.L., Alarcón-Chaidez, F., and Gross, D.C. (1999). *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol Mol Biol Rev *63*, 266–292.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Meth *10*, 563–569.

Clarke, C.R., Cai, R., Studholme, D.J., Guttman, D.S., and Vinatzer, B.A. (2010). *Pseudomonas syringae* strains naturally lacking the classical P. syringae *hrp/hrc* Locus are common leaf colonizers equipped with an atypical type III secretion system. Mol Plant Microbe Interact *23*, 198–210.

Cui, H., Wang, Y., Xue, L., Chu, J., Yan, C., Fu, J., Chen, M., Innes, R.W., and Zhou, J.-M. (2010). *Pseudomonas syringae* Effector Protein AvrB Perturbs Arabidopsis Hormone Signalingby Activating MAP Kinase 4. Cell Host Microbe *7*, 164–175.

Cunnac, S., Chakravarthy, S., Kvitko, B.H., Russell, A.B., Martin, G.B., and Collmer, A. (2011). Genetic disassembly and combinatorial reassembly identify a minimal

functional repertoire of type III effectors in *Pseudomonas syringae*. Proc Natl Acad Sci USA *108*, 2975–2980.

Feil, H., Feil, W.S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., et al. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. Proc Natl Acad Sci USA *102*, 11064–11069.

Guttman, D.S., Vinatzer, B.A., Sarkar, S.F., Ranall, M.V., Kettler, G., and Greenberg, J.T. (2002). A functional screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae*. Science *295*, 1722–1726.

Hirano, S., and Upper, C. (2000). Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*---a pathogen, ice nucleus, and epiphyte. Microbiol Mol Biol Rev *64*, 624.

Hwang, M.S.H., Morgan, R.L., Sarkar, S.F., Wang, P.W., and Guttman, D.S. (2005). Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. Appl Environ Microbiol *71*, 5182–5191.

Jiang, S., Yao, J., Ma, K.-W., Zhou, H., Song, J., He, S.Y., and Ma, W. (2013). Bacterial Effector Activates Jasmonate Signaling by Directly Targeting JAZ Transcriptional Repressors. PLoS Pathog *9*, e1003715.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc *4*, 363–371.

Kvitko, B.H., Park, D.H., Velásquez, A.C., Wei, C.-F., Russell, A.B., Martin, G.B., Schneider, D.J., and Collmer, A. (2009). Deletions in the Repertoire of *Pseudomonas syringae* pv. *tomato* DC3000 Type III Secretion Effector Genes Reveal Functional Overlap among Effectors. PLoS Pathog *5*, e1000388.

Lemonnier, M., Landraud, L., and Lemichez, E. (2007). Rho GTPase-activating bacterial toxins: from bacterial virulence regulation to eukaryotic cell biology. FEMS Microbiol Rev *31*, 515–534.

Lindeberg, M., Cunnac, S., and Collmer, A. (2012). *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. Trends Microbiol *20*, 199–208.

Melotto, M., Underwood, W., Koczan, J., Nomura, K., and He, S.Y. (2006). Plant stomata function in innate immunity against bacterial invasion. Cell *126*, 969-980.

Morris, C.E., Sands, D.C., Vinatzer, B.A., Glaux, C., Guilbaud, C., Buffière, A., Yan, S., Dominguez, H., and Thompson, B.M. (2008). The life history of the plant

pathogen *Pseudomonas syringae* is linked to the water cycle. ISME J *2*, 321-334.

Mucyn TS, Yourstone S, Lind AL, Biswas S, Nishimura MT, **Baltrus DA**, Cumbie JS, Chang JH, Jones CD, Dangl JL, Grant SR *in press* "Variable suites of non-effector genes are co-regulated in the type III secretion virulence regulons across the *Pseudomonas syringae* phylogeny" PLoS Pathogens.

O'Brien, H.E., Thakur, S., and Guttman, D.S. (2011). Evolution of Plant Pathogenesis in *Pseudomonas syringae*: A Genomics Perspective. Annu Rev Phytopathol *49*, 269–289.

Quigley, N.B., and Gross, D.C. (1994). Syringomycin production among strains of *Pseudomonas syringae* pv. *syringae*: conservation of the *syrB* and *syrD* genes and activation of phytotoxin production by plant signal molecules. Mol Plant Microbe Interact *7*, 78–90.

Ramos, C., Matas, I.M., Bardaji, L., Aragón, I.M., and Murillo, J. (2012). *Pseudomonas savastanoi* pv. *savastanoi*: some like it knot. Mol Plant Pathol 13,998-1009

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Systematic Biology *61*, 539–542.

Schellenberg, B., Ramel, C., and Dudler, R. (2010). *Pseudomonas syringae* virulence factor syringolin A counteracts stomatal immunity by proteasome inhibition. Mol Plant Microbe Interact *23*, 1287–1293.

Studholme, D.J. (2011). Application of high-throughput genome sequencing to intrapathovar variation in P*seudomonas syringae.* Mol Plant Pathol 12,829-838

Vinatzer, B.A., Teitzel, G.M., Lee, M.-W., Jelenska, J., Hotton, S., Fairfax, K., Jenrette, J., and Greenberg, J.T. (2006). The type III effector repertoire of *Pseudomonas syringae* pv. *syringae* B728a and its role in survival and disease on host and non-host plants. Mol Microbiol *62*, 26–44.

Acknowledgements

Funding was provided by startup funds to David Baltrus from the University of Arizona.

We thank the University of Delaware Sequencing Core Facility for technical help with

PacBio sequences. We also Jeff Dangl for experimental reagents and positive encouragement.

Figure 1

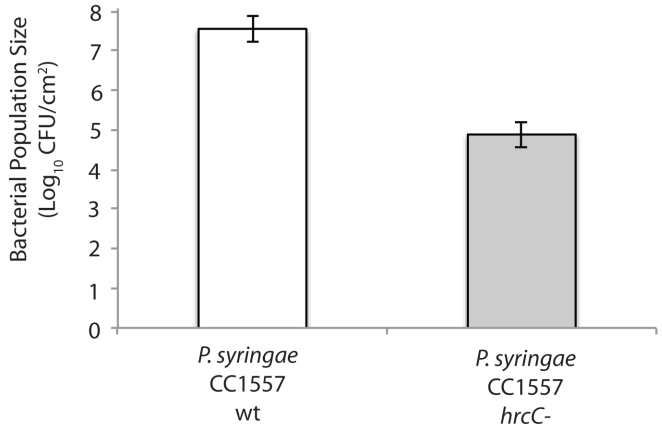

Figure 1. *P. syringae* CC1557 Virulence in *N. benthamiana* is Type III Secretion Dependent. Growth of wild type *P. syringae* CC1557 as well as a *hrcC* mutant was measured *in planta* three days post syringe inoculation. Measurements are based on three independent experiments with at least 4 replicates in each. Bacterial population sizes are significantly different (Tukey's HSD, $p<0.05$) between strains. Error bars represent 2 standard errors.

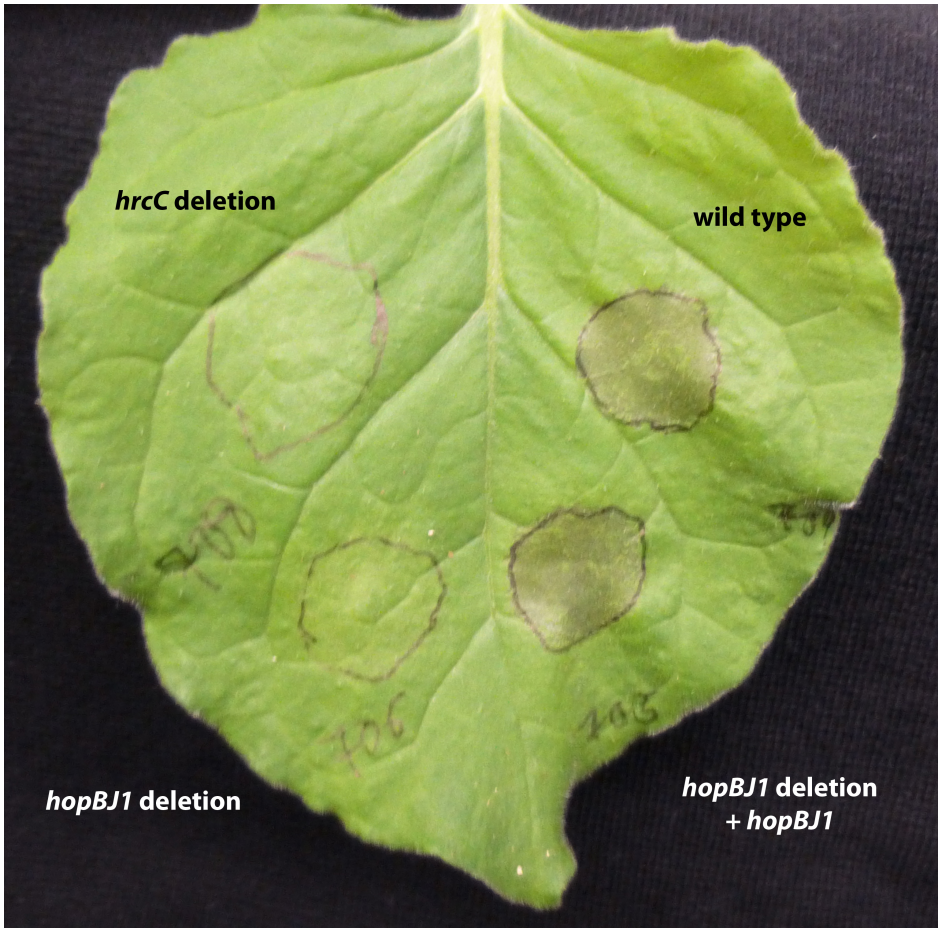

Figure 2. *P. syringae* CC1557 causes disease symptoms within *N. Benthamiana* in a T3SS dependent and HopBJ1 dependent manner. Leaves from two week old *N. Benthamiana* plants were syringe inoculated with various strains derived from *P. syringae* CC1557. Three days after inoculation, plants showed evidence of tissue collapse dependent on a functioning T3SS (compare wild type vs. *hrcC* mutant) as well as HopBJ1 (compare *hopBJ1* deletion strain vs. *hopBJ1* complement). Picture is representative of multiple trials.

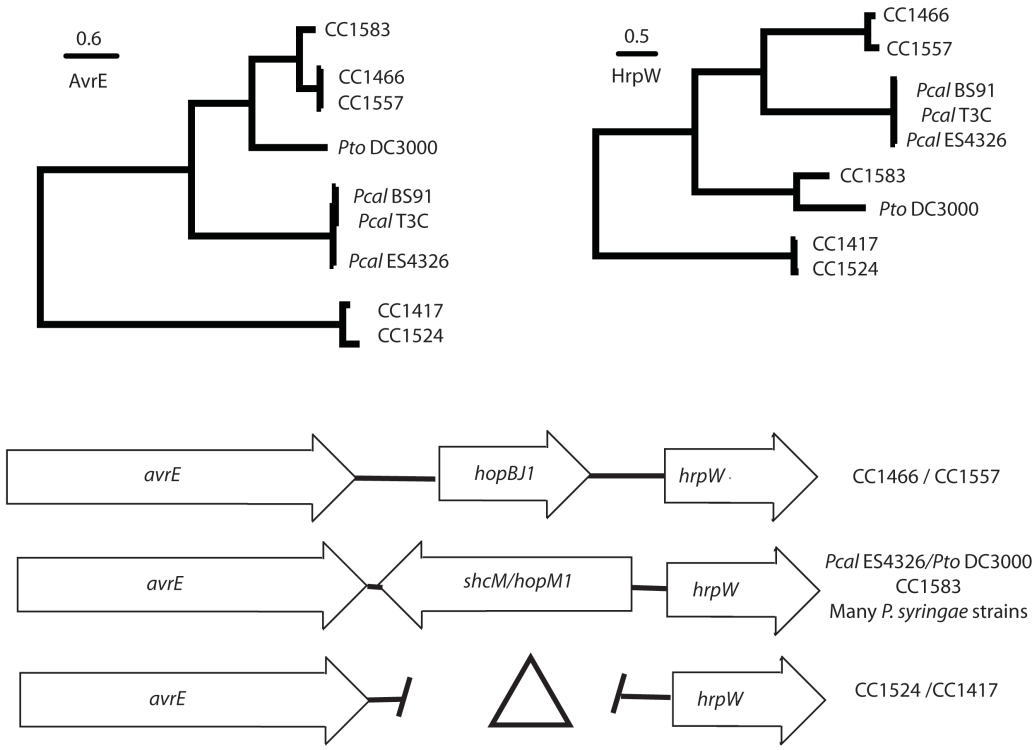

Figure 4

Figure 3. Diversity of the CEL locus across *P. syringae sensu latu*. A) Bayesian phylogenies for AvrE (left) and HrpW (right), for a subset of *P. syringae* strains. Posterior probabilities for all nodes are >0.95. Scale bars indicate number of amino acid changes. Phylogenetic patterns for both of these loci approximate relationships based on core genomes and MLSA loci, except that HrpW from *P. syringae* CC1583 clusters more closely with *P. syringae* pv. *tomato* DC3000. This suggests a recombination event at this locus within *P. syringae* CC1583. **B)** Genomic context of the CEL across *P. syringae sensu latu.* In most strains, *hopM1* is bordered by *avrE* and *hrpW.* Some *P. viridiflava* strains have lost *hopM1*, while CC1557 and CC1466 have replaced this region with *hopBJ1*.

Figure 5

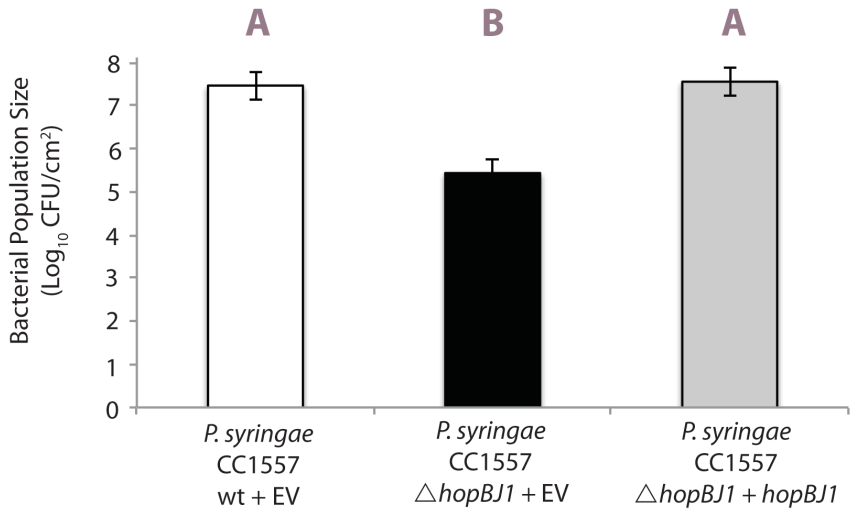

Figure 4. HopBJ1 is a Virulence Factor for *P. syringae* CC1557 in *N. benthamiana*. We created a deletion mutant of *hopBJ1* in *P. syringae* CC1557, and complemented this strain by expressing *hopBJ1* from a plasmid using the native promoter. Deletion of *hopBJ1* lowers growth of *P. syringae* CC1557 three days post inoculation. Both of these phenotypes are complemented by expression of wild type *hopBJ1 in trans*. Three independent experiments were performed with at least 4 replicates in each treatment. Letters represent statistical differences within an ANOVA (Tukey's HSD, $p<0.05$). Error bars represent 2 standard errors.

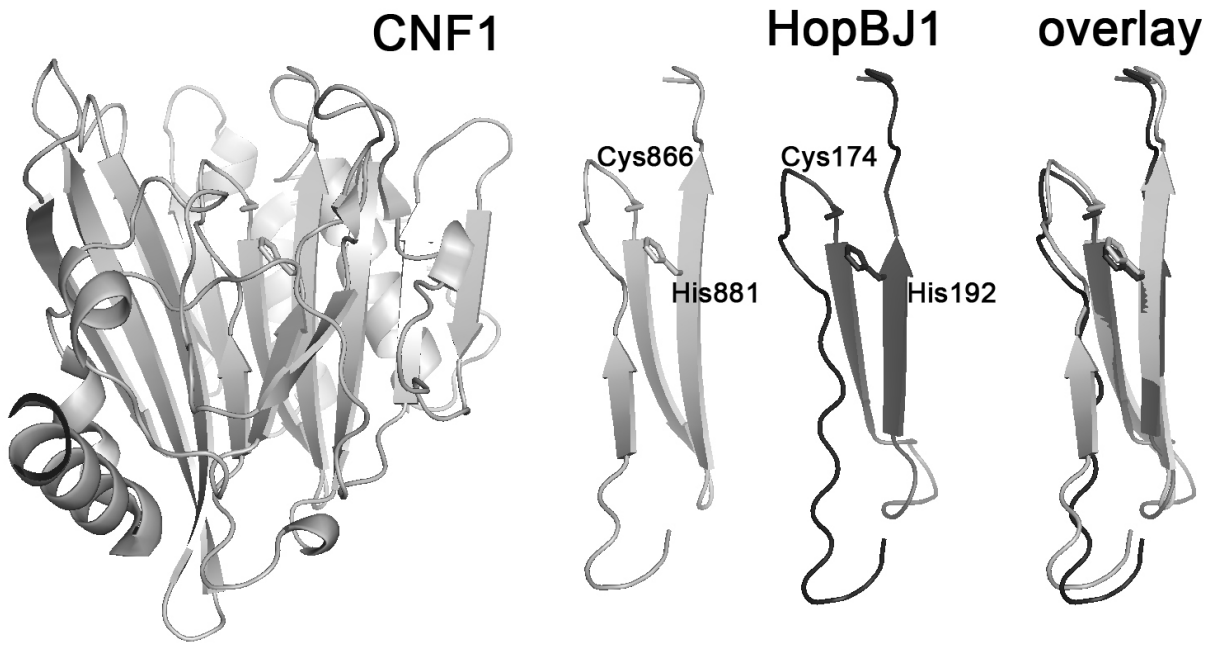

Figure 5. Protein Motifs within HopBJ1 Resemble the *E. coli* CNF1 Toxin. Despite no sequence similarity between HopBJ1 and CNF1 by BLASTp searches, modeling of tertiary protein structure demonstrates limited similarity in protein structure between the two. Importantly, the catalytic dyad of *E. coli* CNF1 (Cys866-His881) perfectly lines up with a putative catalytic dyad within HopBJ1 (Cys174 and His192).

Strain	Description
CC1557	*P. syringae* CC1557
DBL627	A rifampicin resistant version of P. syringae CC1557
DBL704	DBL627 with pDAB334, CC1557 rifR mutant with EV
DBL690	DBL627 with pDBL048 integrated, *hopBJ1* deletion construct integrated into CC1557 rifR
DBL692	DBL627 with pDBL051 integrated, *hrcC* deletion construct integrated into CC1557 rifR
DBL698	DBL627 *hrcC* mutant
DBL696	DBL627 *hopBJ1* mutant
DBL704	DBL627 with EV pDAB334, CC1557 rifR with EV
DBL706	DBL696 del with pDAB334, *hopBJ1* deletion with EV
DBL705	DBL696 del with pDBL052, *hopBJ1* deletion with *hopBJ1*
DBL708	DBL698 with pDAB334, CC1557 rifR *hrcC*- with EV
DBL707	DBL698 with pDBL052, CC1557 rifR *hrcC*- with *hopBJ1*
DBL604	*PtoDC3000 with pDBL061*
DBL888	*PtoDC3000 hrcC mutant with hopBJ1 (no promoter) pDBL061*

Plasmid	Description
MTN1907	Gateway destination vector for making clean deletions in *P. syringae*
pDONR207	Gateway entry vector from Invitrogen
MTN1970-2	the 3' UTR of *Arabidopsis* HSP18.2 in pDONR207
pJC531	Gateway destination vector based off of pBBRMCS1 no promoter
pDAB334	MTN1970-2 EV construct in pJC531
pDBL040	*hopBJ1* cloned with promoter, no stop codon in pDONR207
pDBL052	*hopBJ1* cloned with promoter, no stop codon in pJC531
pDBL042	*hopBJ1* deletion construct in pDONR207
pDBL048	*hopBJ1* deletion construct in MTN1907
pDBL012	*hopBJ1* cloned without promoter into pDONR207, no stop
pDBL061	*hopBJ1* cloned without promter into pBV187
pDBL045	CC1557 *hrcC* deletion construct in pDONR207
pDBL051	CC1557 *hrcC* deletion construct in MTN1907

Table 2. Type III Effector Family Distribution

	CC1557	CC1466	CC1583
avrE	P	P	P
hopF2-1			P
hopF2-2			P
hopM1			P
hopY1	P	P	P
hopAA1			P
hopAG1			P
hopAH1			P
hopAH2	P	P	P
hopAI1			P
hopAS1			P
hopBF1	P	P	P
hopBJ1	P	P	

The letter P denotes presence of this particular effector family, whereas the lack of a letter denotes absence.

Arabidopsis Accession	Reaction
Col	++
Col rps2	++
Ws	++
Ws eds1	++
Ws pad4	++
Ws rar1	++
Ws NahG	++
Ag-0	++
Bur-0	+
Can-0	+
Ct-1	++
Edi-0	++
Hi-0	+
Kn-0	+
Ler-0	++
Mt-0	++
No-0	++
Oy-0	++
Po-0	++
Rsch	++
Sf-0	+
Tsu-0	+
Wil-2	+
Ws-0	+
Wu-0	+
Zu-0	+

++ HR in 24hr
+ HR in 48hr
Gene names next to accession indicate mutant lines

Figure S1. Toxin Regions are Syntenic in *P. syringae* pv. *syringae* B728a and *P. syringae* CC1557. Mauve was used to align four whole *P. syringae* genomes, and the region that contains pathways encoding syringomycin was extracted from *P. syringae* pv. *syringae* B728a. The toxin region for syringomycin and syringopeptin is shown in purple, while adjacent genomic regions conserved throughout the four genomes are listed with either the number 1 or 2. The strain order is (from top): CC1557, pv. *syringae* B728a, pv. *phaseolicola* 1448a, pv. *tomato* DC3000